REMARKS

This is in response to the Office Action mailed on March 12, 2007. Claims 1-23 were pending in the application and the Examiner rejected all claims. With this amendment, claims 1, 5-9, 11, and 18 are amended, claims 2-4, 12-16 and 22-23 are canceled, and the remaining claims are unchanged in the application.

At the top of page 2 of the Office Action, the Examiner objected to claim 9 as depending from itself. The dependency of claim 9 has been changed as suggested by the Examiner, to depend from claim 8. Therefore, Applicant submits that the claims are in proper form.

On page 2 of the Office Action, the Examiner rejected claims 1-9, 11-16 and 18-23 under 35 U.S.C. §102(e) as being anticipated by Basu et al. US Patent No. 6,594,629. Applicant respectfully traverses the Examiner's rejection.

Claim 1 has been amended so that all of the limitations of claims 1-4 are now incorporated into claim 1. Claim 1 therefore claims a speech detector component that detects whether a user is speaking based on a sensor signal output by a speech sensor that senses a non-audio input generated by speech action. The speech detector component outputs the speech detection signal "based on a level of variance in a first characteristic of the sensor signal..., wherein the first characteristic of the sensor signal has a first level of variance when the user is speaking and a second level of variance when the user is not speaking...". The speech detector component "outputs the speech detection signal based on the level of variance of the first characteristic of the sensor signal relative to a baseline level of variance of the first characteristic that comprises a level of a predetermined one of the first and second levels of the characteristic over a given time period."

It can thus be seen that the speech detector in claim 1 monitors the signal level of the signal output by the speech sensor as against a baseline level. The baseline level is determined by monitoring the sensor signal over a period of time, to establish the baseline. It is clear that the baseline level of the first characteristic can be either the level of the first

characteristic when the user is speaking, or when the user is not speaking. This is simply neither taught nor suggested by Basu et al.

In rejecting the limitations of original claims 1- 4, the Examiner cited column 15, lines 37-65, FIG. 1 and FIG. 5 of Basu et al. However, none of these citations either teach or suggest the speech detection system set out in claim 1. The reference simply does not show that speech is detected by comparing a variance level for a non-audio input detector against a baseline variance level. Instead, the portions cited by the Examiner indicate that the video signal, when it is believed that the user is opening his or her mouth, is compared against stored video patterns of the user opening his or her mouth to determine that speech is taking place. This would appear to be much more complex and cumbersome, and require much greater effort in training, than the present system. By contrast, the present system simply takes a baseline measurement of the variance of the non-audio input sensor (such as when the user is speaking or when the user is not speaking) and calculates the baseline variance level. The present system then simply compares the variance level of the sensor input signal against the baseline level to detect speech. Because this is neither taught nor suggested by Basu et al., Applicant submits that independent claim 1 is allowable over Basu et al.

Independent claim 11 has now been amended to include the limitations of original claims 11-14 and 16. Claim 11 is drawn to a speech recognition system that has a speech detector component that "calculates the speech detection signal as a speech detection measure, indicative of a probability that the user is speaking and combines the speech detection measure with the microphone signal to generate a combined signal [which is] a product of the probability and the microphone signal.". The speech recognition engine then recognizes speech in the sensed audio input "based on the combined signal." It is thus clear that the signal that the speech recognizer actually receives and bases recognition on is a product of the probability that the user is speaking (calculated by the speech detector) and the acoustic microphone signal. This is simply neither taught nor suggested by the reference cited by the Examiner.

In order to meet these limitations, the Examiner cited FIG. 5, column 13, lines 36-65, FIG. 8B and column 15, lines 37-65. However, these citations do not teach that the speech

signal upon which speech recognition is performed is a product of the acoustic signal input by the audio microphone and a probability, calculated by a speech detector component, that the speaker is speaking. Instead, at the cited portions of Basu et al., Basu simply operates in one of two modes. The first mode is that Basu uses the video system to detect whether the speaker is speaking. If the detection is positive (that the speaker is speaking) then the microphone is turned on. Otherwise, the microphone is turned off. In the second mode, Basu uses the video system to actually predict visemes (or phonemes) to perform speech recognition. This type of speech recognition is combined with the audio speech recognition in order to generate a recognition result. However, in combining the probabilities generated from the video system with those generated from the audio system, Basu et al. does not teach or suggest that the probability that the speaker is speaking be combined with the probabilities generated from the audio system. Instead, Basu et al. teaches that the probabilities associated with visemes (or phonemes) recognized by the video system are combined with the probabilities of phonemes recognized by the audio system. These are completely different.

In Basu's system, in order to combine probabilities, the video system must actually perform speech recognition, which can be a much more expensive and cumbersome task, then simply computing a speech detection probability. By computing the speech detection probability, instead of speech recognition probabilities, the present system simply needs to compute a probability that the speaker is speaking. This is then multiplied by the signal input by the microphone. In contrast, because Basu et al. calculates speech recognition probabilities with its video system, it must perform significantly more computation.

Independent claim 18 now includes the limitations of original claims 18, 22, and 23. Claim 18 is thus drawn to a method of recognizing speech that includes "detecting whether the user is speaking based on the first and second signals [those from an audio microphone and a facial movement sensor]; and recognizing speech based on the first signal and the speech detection signal, wherein recognizing speech comprises increasing a likelihood that the speech is recognized based on a probability that the speech detection signal indicates that the user is speaking; and decreasing a likelihood that the speech is recognized based on a probability that

the speech detection signal indicates that the speaker is not speaking." It is thus clear that the amount by which the likelihood that speech is recognized is based on the probability that the speaker is speaking, generated by the speech detector. This is simply neither taught nor suggested by Basu et al.

Basu et al. in a first mode either turns on or off the microphone, depending on whether speech is detected. In a second mode, Basu et al. uses the video system to actually recognize visemes (or phonemes). There is no teaching or suggestion that Basu et al. computes a probability that the speaker is speaking and multiplies the speech signal upon which speech recognition is performed, by that probability. Therefore, Applicant submits that independent claim 18 is allowable as well.

In conclusion, Applicant submits that independent claims 1, 11 and 18 are allowable over the references cited by the Examiner. Applicant further submits that dependent claims 5-10, 17, and 19-21, are allowable as well. Reconsideration and allowance of claims 1, 5-11 and 17-21 are respectfully requested.

The Director is authorized to charge any fee deficiency required by this paper or credit any overpayment to Deposit Account No. 23-1123.

Respectfully submitted,

WESTMAN, CHAMPLIN & KELLY, P.A.

By: _____

Joseph R. Kelly, Reg. No. 34,847
900 Second Avenue South, Suite 1400
Minneapolis, Minnesota 55402-3319
Phone: (612) 334-3222 Fax: (612) 334-3312

JRK:slg